**BF[2] - Enrichment**

**Type in your score here 🡪 \_\_\_30\_ out of 30 points possible**

1. (5 points) Ponder/Reflect Exercise – Reflect on what you have learned from this portion of the class. Examples of what you can do are: a brief outline of material covered, insights you gained from class or personal study, or items you feel that you need to follow up or work on. (3-5 sentences)

I learned that Type 1 sum of squares works best on consistent and complete data. Type 3 works when the data is uneven, and the replications have different amounts. This is so you can get the correct p-value and F-statistic.

2. Do a complete ANOVA on cancer file (cancer.txt). Fit an ANOVA model that includes terms for gender (“f” or “m”), cancer type, and the interaction between gender and cancer type.

(a) (5 points) Find the complete ANOVA table USING TYPE I SS. Interpret/state a conclusion for each of the structural factors in the model (i.e. tell whether each term is significant or not). If it is not appropriate to interpret the hypothesis test results for the main effects, state as much and state why.

Analysis of Variance Table

Response: age

Df Sum Sq Mean Sq F value Pr(>F)

type 5 399.8 79.961 0.6629 0.6532

gender 1 50.9 50.877 0.4218 0.5190

type:gender 5 584.1 116.825 0.9685 0.4459

Residuals 51 6151.8 120.623

The p-values are large for the main effects, so we would not reject the null hypothesis. We can say that we have insufficient evidence the means are different between gender as well as type.

The p-value for the interaction is large so we would not reject the null, se say that we have insufficient evidence that we have an interaction.

(b) (5 points) Find the complete ANOVA table USING TYPE III SS. Interpret/state a conclusion for each of the structural factors in the model (i.e. tell whether each term is significant or not). If it is not appropriate to interpret the hypothesis test results for the main effects, state as much and state why.

Anova Table (Type III tests)

Response: age

Sum Sq Df F value Pr(>F)

(Intercept) 9384.5 1 77.7999 7.771e-12 \*\*\*

type 469.0 5 0.7776 0.5705

gender 24.3 1 0.2015 0.6555

type:gender 584.1 5 0.9685 0.4459

Residuals 6151.8 51

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

The p-values are large for the main effects so we would not reject the null hypothesis, we can say we have insufficient evidence the means are different.

The p-value for the interaction is large so we would not reject the null. We have insufficient evidence that we have an interaction.

3. Do a complete analysis of variance on heights of singers in a choir, found in the file singerheights.csv (note that it is comma-delimited). Fit an ANOVA model that includes terms for gender (“f” or “m”), singing part (“low” or “high”), and the interaction between gender and part. (Note that the low part for females is generally called alto, high part for females is soprano, low part for males is bass, and high part for males is tenor. However, we are interested in the association between singing the high/low part and height, so we are treating this as a 2 x 2 factorial instead of a one-way Anova with 4 levels of “singing part.”)

(a) (5 points) Find the complete ANOVA table USING TYPE I SS. Interpret/state a conclusion for each of the structural factors in the model (i.e. tell whether each term is significant or not). If it is not appropriate to interpret the hypothesis test results for the main effects, state as much and state why. For those factors that are significant, explain/show which level of the factor has a higher height and which is associated with a lower height.

Analysis of Variance Table

Response: height

Df Sum Sq Mean Sq F value Pr(>F)

gender 1 1018.86 1018.86 161.1262 < 2e-16 \*\*\*

part 1 33.09 33.09 5.2329 0.02383 \*

gender:part 1 6.58 6.58 1.0413 0.30948

Residuals 126 796.74 6.32

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

The p-values for the main effects are small so we would reject the null. We say we have sufficient evidence the means are different between gender and the part.

The p-value for the interaction is large so we would not reject the null. We say we have insufficient evidence that we have an interaction.

(b) (5 points) Find the complete ANOVA table USING TYPE III SS. Interpret/state a conclusion for each of the structural factors in the model (i.e. tell whether each term is significant or not). If it is not appropriate to interpret the hypothesis test results for the main effects, state as much and state why. For those factors that are significant, explain/show which level of the factor has a higher height and which is associated with a lower height.

Anova Table (Type III tests)

Response: height

Sum Sq Df F value Pr(>F)

(Intercept) 548245 1 86701.8490 <2e-16 \*\*\*

part 37 1 5.8184 0.0173 \*

gender 873 1 138.0042 <2e-16 \*\*\*

part:gender 7 1 1.0413 0.3095

Residuals 797 126

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

The p-values are small for the main effect, so we would reject the null, we have sufficient evidence that the means are different for gender and part.

The p-value for the interaction is large so we would not reject the null and we have insufficient evidence that there is an interaction.

(c) (2 points) Why is the SS for gender so much smaller with Type III SS? Explain.

SS for gender for the type 3 SS is smaller because type 3 SS puts all the other terms in the model, and what is left over is given to the gender SS. In the type 1 SS the first term that is included in the model so there is more sum of squares to work with.

4. (3 pts) What are the main reasons we want to use replication (more than one unit per treatment group)?

Replication gives more precision to our estimate of model parameters. Replications give us more information about the Epsilon which allows us to make inference about the model parameters.